

Review: Times Series

• A time series y_t is a process observed in sequence over time,

 $t=1,...,T \qquad \Rightarrow Y_t=\{y_1,y_2,y_3,...,y_T\}.$

• The main feature of time series: *dependence*.

• With dependent observations, we need new assumptions and tools are needed: stationarity, ergodicity, & CLT for dependenet observations (MDS CLT).

• Roughly speaking, stationarity requires constant moments for Y_t ; ergodicity requires that the dependence is short-lived, eventually y_t has only a small influence on y_{t+k} , when k is relatively large.

Review: Stationarity – Example Example: Assume $\varepsilon_t \sim WN(0, \sigma^2)$. $y_t = \phi y_{t-1} + \varepsilon_t$. (AR(1) process) • Mean $E[y_t] = \mu = 0$ (assuming $\phi \neq 1$) • **Variance** $Var[y_t] = \gamma(0) = \phi^2 Var[y_{t-1}] + Var[\varepsilon_t]$ $\gamma(0) = \sigma^2/(1 - \phi^2)$ (assuming $|\phi| < 1$) • **Covariance** $\gamma(k) = Cov[y_t, y_{t-k}] = \phi^k \gamma(0)$ \Rightarrow If $|\phi| < 1$, AR(1) process is covariance **stationary**. • Auto-correlation function (**ACF**): $\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \phi^k$

Review: Non-Stationarity – Example Example: Assume $\varepsilon_t \sim WN(0, \sigma^2)$. $y_t = \mu + y_{t-1} + \varepsilon_t$ (Random Walk with drift process) Doing backward substitution: $\Rightarrow y_t = \mu t + \sum_{j=0}^{t-1} \varepsilon_{t-j} + y_0$ • Mean & Variance $E[y_t] = \mu t + y_0$ $Var[y_t] = \gamma(0) = \sum_{j=0}^{t-1} \sigma^2 = \sigma^2 t$ \Rightarrow the RW process is non-stationary; that is, moments are time dependent.

Review: Stationarity: Remarks

• Stationarity is an invariant property: The statistical characteristics of the time series do not vary over time.

• If IBM is weak stationary, then, the returns of IBM may change month to month or year to year, but the average return and the variance in two equal lengths time intervals will be more or less the same.

• In the long run, say 100-200 years, the stationarity assumption may not be realistic.

• In general, time series analysis is done under the stationarity assumption.

Review: Ergodicity

• We want to estimate the mean of the process $\{Z_t\}$, $\mu(Z_t)$. But, we need to distinguishing between *ensemble average* (with *m cross section* observations) and *time average* (with *T time series* observations):

- Ensemble Average: $\overline{\overline{z}} = \frac{\sum_{i=1}^{m} Z_i}{m}$

- Time Series Average: $\overline{z} = \frac{\sum_{t=1}^{T} z_t}{T}$

Q: Which estimator is the most appropriate? A: Ensemble Average, \overline{z} . But, we cannot compute it for a time series.

• The *Ergodic Theorem* tells us when \overline{z} can be used instead of \overline{z} .

Theorem: A sufficient condition for ergodicity for the mean:

 $\rho_k \to 0$ as $k = t_i - t_j \to \infty$

We need the correlation between (y_{t_i}, y_{t_j}) to decrease as they grow further apart in time.

Review: Invertibility

• Invertibility allows us to convert an MA process into an AR process. AR processes are easier to use and estimate.

Example: Suppose we have an MA(1) process:

$$y_t = \mu + \theta_1 \varepsilon_{t-1} + \varepsilon_t = \mu + \theta(L)\varepsilon_t - \theta(L) = (1 + \theta_1 L)$$

Now, we multiply $\theta(L)^{-1}$ on both sides of the MA process $y_t = \mu + \theta(L) \varepsilon_t$.

$$\Rightarrow \qquad \theta(L)^{-1} y_t = \theta(L)^{-1} \mu + \varepsilon_t = \mu^* + \varepsilon_t$$

Then, we get an $AR(\infty)$:

$$y_t = \mu * + \theta_1 y_{t-1} - \theta_1^2 y_{t-2} + \theta_1^3 y_{t-3} - \theta_1^4 y_{t-4} + \dots + \varepsilon_t$$

If the resulting AR(∞) process is non-explosive, then, the MA(1) is **invertible**. The invertibility condition in this case: $|\theta_1| < 1$.

Review: Moving Average Process

• An MA process models $E_t[y_t | I_{t-1}]$ with lagged error terms. An MA(q) model involves q lags of ε_t .

• We keep $\varepsilon_t \sim WN(0, \sigma^2)$

Example: A linear MA(q) model:

 $y_t = \mu + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t = \mu + \theta(L) \varepsilon_t,$

where

$$\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \theta_2 L^3 + \dots + \theta_q L^q$$

• In time series, the constant does not affect the properties of AR and MA process. Thus, in this situation we say "without loss of generalization", we assume $\mu = 0$.

Review: MA(q) Process– Stationarity				
• Q: Is MA(q) stationary? Check the moments (assume $\mu = 0$). $y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + + \theta_q \varepsilon_{t-q}$				
• Mean $E[y_t] = E[\varepsilon_t] + \theta_1 E[\varepsilon_{t-1}] + \theta_2 E[\varepsilon_{t-2}]$	$[\mathbf{r}_2] + + \mathbf{\theta}_q \mathbf{E}[\mathbf{\varepsilon}_{t-q}] = 0$			
• Variance $Var[y_t] = Var[\varepsilon_t] + \theta_1^2 Var[\varepsilon_{t-1}] + \theta_2^2 Var[\varepsilon_{t-2}] + \dots + \theta_q^2 Var[\varepsilon_{t-q}]$ $= (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) \sigma^2.$				
• Covariance For the <i>k</i> autocovariance: $\gamma(k) = \sigma^2 \sum_{j=k}^{q} \theta_j \ \theta_{j-k}$ $\gamma(k) = 0$	for $ k \le q$ $(\theta_0 = 1)$ for $ k > q$			
<u>Remark:</u> After lag q , the autocovariances	(& ACFs) are 0.			



Review: Moving Average Process – Stationarity

• It is easy to verify that the sums $\sum_{j=k}^{q} \theta_j \theta_{j-k}$ are finite. Then, mean, variance and covariance are constant.

 \Rightarrow MA(q) is always stationary.

• <u>Problem</u>: It can be shown that for ε_t with same distribution (say, normal) the autocovariances are non-unique. Then, we select the **invertible** model, with an AR(∞) (non-explosive) representation.

<u>Technical note</u>: An invertible MA(q) is typically required to have roots of the lag polynomial equation $\theta(z) = 0$ greater than one in absolute value (**outside the unit circle**). In the MA(1) case, we require $|\theta_1| < 1$

Review: MA Process – ACF • Recall the **autocorrelation function** (**ACF**): $\rho(k) = \gamma(k)/\gamma(0)$ • Then, for an MA(q) process, the **ACF**: $\rho(k) = \frac{\sum_{j=k}^{q} \theta_{j} \theta_{j-k}}{(1 + \theta_{1}^{2} + \theta_{2}^{2} + ... + \theta_{q}^{2})}$ for $|k| \le q$ $\rho(k) = 0$ for |k| > q<u>Remark</u>: After lag q, the ACF are 0 (contrast with AR(1) model). <u>Note</u>: The ACF is usually shown in a plot. When we plot $\rho(k)$ against k, we plot also $\rho(0) = 1$. • The sample **correlogram** is the plot of the ACF against k. As the ACF lies between -1 and +1, the correlogram also lies between these values.

Review: MA Process – ACF for MA(1)

Autocorrelations

$$\begin{split} \rho(0) &= \gamma(0)/\gamma(0) = 1\\ \rho(1) &= \gamma(1)/\gamma(0) = \theta_1/(1+\theta_1^2)\\ \rho(2) &= \gamma(2)/\gamma(0) = 0\\ \vdots\\ \rho(k) &= \gamma(k)/\gamma(0) = 0 \quad (\text{for } k > 1) \end{split}$$
Note that $|\rho(1)| \leq 0.5$. When $\theta_1 = 0.5 \implies \rho(1) = 0.4$. $(|\theta_1| < 1 \implies \text{invertible})\\ \theta_1 = -0.9 \implies \rho(1) = -0.497238$. $(|\theta_1| < 1 \implies \text{invertible})\\ \theta_1 = 2 \implies \rho(1) = 0.4$. $(|\theta_1| > 1 \text{ non-invertible})$ Note: We have two processes, with the same ACF, we select $\theta_1 = 0.5$.



Review: MA Process – Estimation

• MA processes are more complicated to estimate since we do not observe the errors, ε_t 's: Direct estimation is impossible.

• Two indirect ways:

(1) Using method of moments (MM): We match observed moments and solved for the parameters. For example, for an MA(1):

$$\rho(1) = \theta_1 / (1 + \theta_1^2)$$

$$r_1 = \frac{\hat{\theta}}{(1 + \hat{\theta}^2)} \quad \Rightarrow \qquad \hat{\theta} = \frac{1 \pm \sqrt{1 - 4r_1^2}}{2r_1}$$

• A nonlinear solution and difficult to solve.

(2) Using AR(∞) representation: For MA(1) & $|\theta| < 1$, find $a \in (-1; 1)$ $\varepsilon_t(a) = y_t + a y_{t-1} + a^2 y_{t-2} + a^3 y_{t-3} + \dots$

and look (numerically) for the least-square estimator

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left\{ \mathrm{S}(\boldsymbol{y}; \boldsymbol{\theta}) = \sum_{t=1}^{T} \varepsilon_t(\boldsymbol{a})^2 \right\} \qquad (\boldsymbol{a}^t = \boldsymbol{\theta}_1^t.)$$

Review: Autoregressive (AR) Process

• We model the conditional expectation of y_t , $E_t[y_t | I_{t-1}]$, as a function of its past history.

• We keep
$$\varepsilon_t \sim WN(0, \sigma^2)$$
.

• The most common models are AR models. An AR(1) model involves a single lag, while an AR(*p*) model involves *p* lags. Then, the AR(*p*) process is given by:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim WN.$$

Using the lag operator we write the AR(p) process: $\phi(L) y_t = \varepsilon_t$

 $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$

with

Review: AR Process – AR(1): Stability

• We can analyze the stability from the point of view of the roots of the lag polynomial. For the AR(1) process

$$\phi(z) = 1 - \phi_1 z = 0 \implies |z| = \frac{1}{|\phi_1|} > 1$$

That is, the AR(1) process is stable if the root of $\phi(z)$ is greater than one (also said as "the roots lie outside the unit circle").

This result generalizes to AR(p) process.





Review: AR(1) Process – ACF • An AR(1) model: $y_t = \phi_1 y_{t-1} + \varepsilon_t$, $\varepsilon_t \sim WN$. Last lecture, under the stationarity condition $|\phi_1| < 1$, we derived: • Moments $E[y_t] = \mu = 0$ (assuming $\phi_1 \neq 1$) $Var[y_t] = \gamma(0) = \sigma^2/(1 - \phi_1^2)$ (assuming $|\phi_1| < 1$) $\gamma(k) = \phi_1^k \gamma(0)$ • We also derived the ACF: $\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \frac{\phi_1^k \gamma(0)}{\gamma(0)} = \phi_1^k$ Remark: Assuming $|\phi_1| < 1$, the ACF decays with k.

Review: AR(1) Process – ACF

• ACF for an AR(1) process:

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \frac{\phi_1^k \gamma(0)}{\gamma(0)} = \phi_1^k$$

Note: The plot of $\rho(k)$ against k, is called **autocorrelogram**. We plot also $\rho(0)$, which is 1.

Note:

$- \text{when} 0 < \phi_1 < 1$	\Rightarrow All autocorrelations are positive.
$- \text{when } -1 < \phi_1 < 0$	\Rightarrow The sign of $\rho(k)$ shows an alternating sign
	pattern beginning with a negative value.
– when $\phi_1 = 1$	\Rightarrow AR(1) is non-stationary, $\rho(k) = 1$, for all k .
	Present & past are always correlated!













AR(2) Process – Stationarity & ACF • An AR(2) model: $y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$, $\varepsilon_t \sim WN$. • Moments: $(\mu = 0)$ $E[y_t] = \frac{\mu}{(1 - \phi_1 - \phi_2)} = 0$ (assuming $\phi_1 + \phi_2 \neq 1$) $Var[y_t] = \frac{\sigma^2}{(1 - \phi_1^2 - \phi_2^2)}$ (assuming $\phi_1^2 + \phi_2^2 < 1$) • Autocovariance function $\gamma(k) = Cov[y_t, y_{t-k}] = E[(\phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t) y_{t-k}]$ $= \phi_1 E[y_{t-1} y_{t-k}] + \phi_2 E[y_{t-2} y_{t-k}] + E[\varepsilon_t y_{t-k}]$ $= \phi_1 \gamma(k - 1) + \phi_2 \gamma(k - 2) + E[\varepsilon_t y_{t-k}]$ We have a recursive formula.

AR(2) Process – Stationarity & ACF • Recursive formula: $\gamma(k) = \phi_1 \gamma(k-1) + \phi_2 \gamma(k-2) + E[\varepsilon_t y_{t-k}]$ $(k=0) \quad \gamma(0) = \phi_1 \gamma(-1) + \phi_2 \gamma(-2) + E[\varepsilon_t y_t]$ $= \phi_1 \gamma(1) + \phi_2 \gamma(2) + \sigma^2$ $(k=1) \quad \gamma(1) = \phi_1 \gamma(0) + \phi_2 \gamma(1) + E[\varepsilon_t y_{t-1}]$ $= \phi_1 \gamma(0) + \phi_2 \gamma(1) + 0$ $\Rightarrow \gamma(1) = [\phi_1/(1-\phi_2)] \gamma(0)$ $(k=2) \quad \gamma(2) = \phi_1 \gamma(1) + \phi_2 \gamma(0) + E[\varepsilon_t y_{t-2}]$ $= \phi_1 \gamma(1) + \phi_2 \gamma(0) + 0$ $\Rightarrow \gamma(2) = [\phi_1^2/(1-\phi_2) + \phi_2] \gamma(0)$ Replacing $\gamma(1)$ and $\gamma(2)$ back to $\gamma(0)$: $\gamma(0) = [\phi_1^2/(1-\phi_2)] \gamma(0) + [\phi_2 \phi_1^2/(1-\phi_2) + \phi_2^2] \gamma(0) + \sigma^2$ $= \frac{\sigma^2(1-\phi_2)}{(1-\phi_2) - \phi_1^2(1+\phi_2) + \phi_2^2(1-\phi_2)} \Rightarrow |\phi_2| < 1$

AR(2) Process – Stationarity & ACF • Dividing the recursive formula for $\gamma(k)$ by $\gamma(0)$, we get the ACF: $\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \phi_1 \rho(k-1) + \phi_2 \rho(k-2) + \frac{E[\varepsilon_t y_{t-k}]}{\gamma(0)}$ (k=0) $\rho(0) = 1$ (k=1) $\rho(1) = \phi_1/(1-\phi_2)$ (k=2) $\rho(2) = \phi_1 \rho(1) + \phi_2 \rho(0) = \phi_1^2/(1-\phi_2) + \phi_2$ (k=3) $\rho(3) = \phi_1 \rho(2) + \phi_2 \rho(1) = = \phi_1^3/(1-\phi_2) + \phi_1 \phi_2 + \phi_2 \phi_1/(1-\phi_2)$ Remark: Again, we see exponential decay in the ACF. From the work above, for stationarity, we need: $\phi_1 + \phi_2 \neq 1$. $\phi_1^2 + \phi_2^2 < 1$. $|\phi_2| < 1$.

AR Process – Stationarity and Ergodicity

Theorem: The linear AR(p) process is strictly stationary and ergodic if and only if the roots of $\phi(L)$ are $|z_j| > 1$ for all j, where $|z_j|$ is the modulus of the complex number r_j .

<u>Note</u>: If one of the z_j 's equals 1, $\phi(L)$ (& y_t) has a **unit root** –i.e., $\phi(1)=0$. This is a special case of *non-stationarity*.

• Recall $\phi(L)^{-1}$ produces an infinite sum on the ε_{t-j} 's. If this sum does not explode, we say the process is **stable**.

AR Process – Estimation and Properties

• We go back to the general AR(p). Define

$$\mathbf{x}_t = (1 \ y_{t-1} \ y_{t-2} \dots y_{t-p})$$
$$\mathbf{\beta} = (\mu \ \phi_1 \ \phi_2 \ \dots \ \phi_p)$$

Then the model can be written as

$$y_t = \mathbf{x}_t' \mathbf{\beta} + \varepsilon_t$$

• The OLS estimator is $\mathbf{b} = (X'X)^{-1}X'y$

• Properties:

- Using the Ergodic Theorem, OLS estimator is consistent.
- Using the MDS CLT, OLS estimator is asymptotically normal.

 \Rightarrow asymptotic inference is the same.

• The asymptotic covariance matrix is estimated just as in the crosssection case: The sandwich estimator.

ARMA Process

• A combination of AR(*p*) and MA(*q*) processes produces an ARMA(*p*, *q*) process: $y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_n y_{t-n}$

$$+ \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$
$$= \mu + \sum_{i=1}^p \phi_i y_{t-i} - \sum_{i=1}^q \theta_i L^i \varepsilon_t + \varepsilon_t$$
$$\Rightarrow \phi(L) y_t = \mu + \theta(L) \varepsilon_t$$

• Usually, we insist that $\phi(L) \neq 0$, $\theta(L) \neq 0$ & that the polynomials $\phi(L)$, $\theta(L)$ have no common factors. This implies it is not a lower order ARMA model.

• Stationarity conditions: Since MA(q) processes are always stationary, the stationarity conditions come from the AR(p) part. Thus, we require the roots $\phi(L) = 0$ to be **outside the unit circle**.

ARMA Process – Common Factors

An ARMA(p, q) model with common factors has a lower order ARMA model. That is, a lower p and q.

Example: Common factors. Suppose we have the following ARMA(2, 3) model $y_t = 0.6 y_{t-1} - 0.3 y_{t-2} + \varepsilon_t - 1.4 \varepsilon_{t-1} + 0.9 \varepsilon_{t-2} + 0.3 \varepsilon_{t-3}$ with $\phi(L) = 1 - .6L + .3L^2$ $\theta(L) = 1 - 1.4L + .9L^2 - .3L^3 = (1 - .6L + .3L^2)(1 - L)$ This model simplifies to: $y_t = (1 - L)\varepsilon_t$ $= \varepsilon_t - \varepsilon_{t-1} \implies$ an MA(1) process. • Simplify the common factors and keep the simpler representation.

ARMA Process – Representation

• ARMA(
$$p, q$$
) model:
 $\phi(L)(y_t - \mu) = \theta(L)\varepsilon_t$
• Cases:
Pure AR Representation: $\Pi(L)(y_t - \mu) = \varepsilon_t \Rightarrow \Pi(L) = \frac{\phi_p(L)}{\theta_q(L)}$
Pure MA Representation: $(y_t - \mu) = \Psi(L)\varepsilon_t \Rightarrow \Psi(L) = \frac{\theta_q(L)}{\phi_p(L)}$
Special cases: $-p = 0$: MA(q)
 $-q = 0$: AR(p).



ARMA(1, 1) – Stationarity & ACF • We have a recursive formula: $\gamma(k) = \phi_1 \gamma(k-1) + E[\varepsilon_t y_{t-k}] + \theta_1 E[\varepsilon_{t-1} y_{t-k}]$ It can be shown, after a lot of algebra: For k = 0, $\gamma(0) = \phi_1 \gamma(1) + \sigma^2 + \theta_1 (\phi_1 \sigma^2 + \theta_1 \sigma^2)$ For k = 1, $\gamma(1) = \phi_1 \gamma(0) + \theta_1 \gamma(1)$ For k = 2, $\gamma(2) = \phi_1 \gamma(1)$ For k, $\gamma(k) = \phi_1^{k-1} \gamma(1), \quad k > 1$ $\Rightarrow \text{ If } |\phi_1| < 1, \text{ exponential decay.}$

ARMA(1, 1) – Stationarity & ACF • Two equations for $\gamma(0)$ and $\gamma(1)$: $\gamma(0) = \phi_1 \gamma(1) + \sigma^2 + \theta_1 (\phi_1 \sigma^2 + \theta_1 \sigma^2)$ $\gamma(1) = \phi_1 \gamma(0) + \theta_1 \gamma(1)$ Solving for $\gamma(0) \& \gamma(1)$: $\gamma(0) = \sigma^2 \frac{1 + \theta_1^2 + 2 \phi_1 \theta_1}{1 - \phi_1^2}$ $\gamma(1) = \sigma^2 \frac{(1 + \phi_1 \theta_1) * (\phi_1 + \theta_1)}{1 - \phi_1^2}$: $\gamma(k) = \phi_1^{k-1} \gamma(1), \quad k > 1 \quad \Rightarrow \text{ If } |\phi_1| < 1, \text{ exponential decay.}$ Note: If stationary, ARMA(1,1) & AR(1) show exponential decay.

ARMA: Stationarity, Causality and Invertibility

Theorem: If $\phi(L)$ and $\theta(L)$ have no common factors, a (unique) *stationary* solution to $\phi(L)y_t = \theta(L)\varepsilon_t$ exists if and only if

 $|z| \le 1 \Rightarrow \phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p \ne 0.$

(i.e., roots of $\phi(z) = 0$ need to be **outside the unit circle**, |z| > 1.)

This ARMA(p, q) model is causal if and only if $|z| \le 1 \Rightarrow \phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p \neq 0.$

This ARMA(p, q) model is invertible if and only if

$$|z| \le 1 \Rightarrow \theta(z) = 1 + \theta_1 z - \theta_2 z^2 + \ldots + \theta_p z^p \neq 0.$$

<u>Note</u>: Real data cannot be *exactly* modeled using a finite number of parameters. We choose p, q to create a good approximated model.

ARMA Process: Identification and Estimation

• We defined the ARMA(p, q) model: $\phi(L)(y_t - \mu) = \theta(L)\varepsilon_t$

The mean does not affect the order of the ARMA. Then, if $\mu \neq 0$, we demean the data: $x_t = y_t - \mu$.

Then, $\phi(L) x_t = \theta(L) \varepsilon_t \implies x_t$ is a demeaned ARMA process.

• For the rest of the lecture, we will study:

- Identification of *p*, *q*.

- Estimation of ARMA(p, q)

ACF: Estimation (System of Equations) • The pxp system of equations: $\gamma(1) = E[y_t, y_{t-1}] = \phi_1 \gamma(0) + \phi_2 \gamma(1) + \phi_3 \gamma(2) + \dots + \phi_p \gamma(p-1)$ $\gamma(2) = E[y_t, y_{t-2}] = \phi_1 \gamma(1) + \phi_2 \gamma(0) + \phi_3 \gamma(1) + \dots + \phi_p \gamma(p-2)$ $\gamma(3) = E[y_t, y_{t-3}] = \phi_1 \gamma(2) + \phi_2 \gamma(1) + \phi_3 \gamma(0) + \dots + \phi_p \gamma(p-3)$ $\vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots$ Using linear algebra, we write the system as: $\gamma = \Gamma \phi$ where $\Gamma = \begin{bmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(p-2) \\ \vdots & \vdots & \vdots & \vdots \\ \gamma(p-1) & \gamma(p-2) & \cdots & \gamma(0) \end{bmatrix}$ a pxp matrix ϕ is the px1 vector of AR(p) coefficients γ is the px1 vector of $\gamma(k)$ autocovariances.

ACF: Estimation – Yule-Walker

• Now, we define the autocorrelation function (**ACF**): $\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \frac{\text{covariance at lag } k}{\text{variance}}$ The ACF lies between -1 and +1, with $\rho(0) = 1$. • Dividing the autocovariance system by $\gamma(0)$, we get: $\begin{bmatrix} \rho(0) & \rho(1) & \cdots & \rho(p-1) \\ \rho(1) & \rho(0) & \cdots & \rho(p-2) \\ \vdots & \vdots & \cdots & \vdots \\ \rho(p-1) & \rho(p-2) & \cdots & \rho(0) \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} = \begin{bmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(p) \end{bmatrix}$ Or using linear algebra: $\mathbf{P} \phi = \rho$ • These are **Yule-Walker** equations, which can be solved numerically.

ACF: Estimation & Correlogram

• Estimation:

Easy: Use sample moments to estimate $\gamma(k)$ and plug in formula:

$$r_k = \hat{\rho}_k = \frac{\sum (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{\sum (Y_t - \bar{Y})^2}$$

We plug $\hat{\rho}_k = r_k$ in the Yule-Walker equations and solve for ϕ :

$$\boldsymbol{R} \boldsymbol{\phi} = \boldsymbol{r} \qquad \Rightarrow \boldsymbol{\phi} = \boldsymbol{R}^{-1} \boldsymbol{r}$$

where \boldsymbol{R} is the estimated correlation matrix \boldsymbol{P} .

• The sample *correlogram* is the plot of the ACF against k. As the ACF lies between -1 and +1, the correlogram also lies between these values.

ACF – Distribution

• Distribution:

For a linear, stationary process, with large *T*, the distribution of the sample ACF, $r_k = \hat{\rho}_k$ is approximately normal with:

 $\mathbf{r} \longrightarrow \mathcal{N}(\mathbf{\rho}, \mathbf{V}/T),$ **V** is the covariance matrix.

Under H_0 (no autocorrelations) $\rho_k = 0$ for all k > 1.

$$\mathbf{r} \xrightarrow{u} N(\mathbf{0}, \mathbf{I}/T) \implies \operatorname{Var}[r_k] = 1/T.$$

• Under H₀, the SE = $1/\sqrt{T}$

$$\Rightarrow$$
 95% C.I.: 0 \pm 1.96 * 1/ \sqrt{T}

Then, for an uncorrelated, WN sequence, approximately 95% of the sample ACFs should be within the above C.I. limits.

<u>Note</u>: The SE = $1/\sqrt{T}$ are sometimes referred as *Bartlett's SE*.

ACF – Identification

• The ACF can be used as a tool to select an ARMA(p, q) model. In general, it is used to select the lag q in an MA model.

	AR(p)	MA(q)	ARMA(p, q)
ACF	Tails off	0 after lag q	Tails off

<u>Note</u>: Ideally, "Tails off" is exponential decay. In practice, we may see decay with a lot of "noise" and a lot of non-zero values.

• In the next slides, we simulate ARMA models. This is an "ideal" situation, we know the model that generated the data. Then, we look at the ACF to see if it is easy to guess the model and order of the model.

ACF - AR(1)

Example: Sample ACF for an AR(1) process:

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \phi_1^k \qquad k = 0, 1, 2, \dots$$

Under stationarity, $|\phi_1| \le 1$, the ACF will show exponential decay.

• Suppose $\phi_1 = 0.4$. Then,

 $\rho(0) = 1$ $\rho(1) = 0.4$ $\rho(2) = 0.4^2 = 0.16$ $\rho(3) = 0.4^3 = 0.064$ $\rho(4) = 0.4^4 = 0.0256$ $\rho(k) = 0.4^k$

÷

ACF – AR(1) Example (continuation): $\rho(k) = 0.4^k$ We simulate an AR(1) series with with $\phi_1 = 0.4$, using the R function *arima.sim*. $\sin_ar1_04 <- arima.sim(list(order=c(1,0,0), ar=0.4), n=20)$ #simulate AR(1) series plot(sim_ar1_04, ylab="Simulated Series", main=(expression(AR(1):~~~phi==0.4))) $acf(sim_ar1_04)$ #plot ACF for sim series Recall SE = $1/\sqrt{T} = 1/\sqrt{200} = 0.07071068$ $\Rightarrow 95\%$ C.I.: $0 \pm 1.96 * 0.07071068 = [-0.1386, 0.1386]$



ACF - MA(1) **Example (continuation)**: Sample ACF for an MA(1) process. $\rho(0) = 1$ $\rho(1) = \theta_1 / (1 + \theta_1^2)$ for k = 1, -1 $\rho(k) = 0$ for |k| > 1. After k = 1 –i.e., one lag– the ACF dies out. Suppose $\theta_1 = 0.5$. Then, $\rho(0) = 1$ $\rho(1) = 0.4$ $\rho(k) = \mathbf{0}$ for |k| > 1. We simulate an MA(1) series with $\phi_1 = 0.4$ sim_ma1_05 <- arima.sim(list(order=c(0,0,1), ma=0.5), n=200) #simulate MA(1) series plot(sim_ma1_05, ylab="Simulated Series", main=(expression(MA(1):~~~theta==0.5))) acf(sim_ma1_05) #plot ACF for sim series



ACF - MA(q)Example: Sample ACF for an MA(q) process: $y_t = \mu + \varepsilon_t + \theta_1 \ \varepsilon_{t-1} + \theta_2 \ \varepsilon_{t-2} + \dots + \theta_q \ \varepsilon_{t-q}$ $\rho(k) = \frac{\sum_{j=k}^q \theta_j \theta_{j-k}}{(1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2)} \qquad k \le q$ $= 0 \qquad \text{otherwise.}$ For different k's: $\rho(0) = 1$ $\rho(1) = \frac{\theta_1 + \theta_2 \theta_1 + \theta_3 \theta_2}{(1 + \theta_1^2 + \theta_2^2 + \theta_3^2)}$ $\rho(2) = \frac{\theta_2 + \theta_3 \theta_1}{(1 + \theta_1^2 + \theta_2^2 + \theta_3^2)}$ $\rho(3) = \frac{\theta_3}{(1 + \theta_1^2 + \theta_2^2 + \theta_3^2)}$ $\rho(k) = 0 \qquad \text{for } |k| > 3.$





27

ACF - ARMA(1, 1)

Example : Sample ACF for an ARMA(1,1) process:
$y_t = \phi_1 y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$
• From the autocovariances, we get
$\gamma(0) = \sigma^2 \frac{1 + \theta_1^2 + 2\phi_1 \theta_1}{1 - \phi_1^2}$
$\gamma(1) = \sigma^2 \frac{(1 + \phi_1 \theta_1) * (\phi_1 + \theta_1)}{1 - \phi_1^2}$
$\gamma(k) = \phi_1 \gamma(k-1) = \phi_1^{k-1} \sigma^2 \frac{(1+\phi_1 \theta_1) * (\phi_1 + \theta_1)}{1-\phi_1^2}$
• Then,
$\rho(k) = \phi_1^{k-1} \frac{(1+\phi_1\theta_1)*(\phi_1+\theta_1)}{1+\theta_1^2+2\phi_1\theta_1}$
\Rightarrow If $ \phi_1 < 1$, exponential decay. Similar pattern to AR(1).

ACF - ARMA(1,1)

Example (continuation): Sample ACF for an ARMA(1,1) process: $y_t = \phi_1 y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$ The ACF for an ARMA(1,1): $\rho(k) = \phi_1^{k-1} \frac{(1+\phi_1\theta_1)*(\phi_1+\theta_1)}{1+\theta_1^2+2\phi_1\theta_1}$ • Suppose $\phi_1 = 0.4, \theta_1 = 0.5$. Then, $\rho(0) = 1$ $\rho(1) = \frac{(1+0.4*0.5)*(0.4+0.5)}{1+0.5^2+2*0.4*0.5} = 0.6545$ $\rho(2) = 0.4*\frac{(1+0.4*0.5)*(0.4+0.5)}{1+0.5^2+2*0.4*0.5} = 0.2618$ $\rho(3) = 0.4^2*\frac{(1+0.4*0.5)*(0.4+0.5)}{1+0.5^2+2*0.4*0.5} = 0.0233$: $\rho(k) = 0.4^{k-1}*\frac{(1+0.4*0.5)*(0.4+0.5)}{1+0.5^2+2*0.4*0.5}$











ACF – Joint Significance Tests

• The application of the LB test to the ACF is straightforward. We use the Ljung-Box (LB) statistic to test $H_0: \rho_1 = \rho_2 = ... = \rho_m = 0$.

Under H₀:

$$LB = T(T+2) \sum_{k=1}^{m} \left(\frac{\hat{\rho}_k^2}{(T-k)}\right) \xrightarrow{d} \chi_m^2$$

Example: LB test with **20 lags** for **US Monthly Returns** and **Changes in Dividends** (1871 – 2020)

> Box.test(lr_p, lag=20, type= "Ljung-Box") data: lr_p X-squared = 208.02, df = 20, p-value < 2.2e-16 \Rightarrow Reject H₀ at 5% level. > Box.test(lr_d, lag=20, type= "Ljung-Box") data: lr_d X-squared = 2762.7, df = 20, p-value < 2.2e-16 \Rightarrow Reject H₀ at 5% level. <u>Conclusion</u>: We found joint significance of first 20 autocorrelations.

Partial ACF (PACF)

• The ACF gives us a lot of information about the order of the dependence when the series we analyze follows a MA process: The ACF is zero after q lags for an MA(q) process.

• If the series we analyze, however, follows an ARMA or AR, the ACF alone tells us little about the orders of dependence: We only observe an exponential decay.

• We introduce a new function that behaves like the ACF of MA models, but for AR models, namely, the partial autocorrelation function (PACF).

• The PACF is similar to the ACF. It measures correlation between observations that are k time periods apart, after controlling for correlations at intermediate lags.

Partial ACF Intuition: Suppose we have an AR(1): $y_t = \phi_1 y_{t-1} + \varepsilon_t$. Then, $\rho(2) = \phi_1^2$ The correlation between y_t and y_{t-2} is not zero, as it would be for an MA(1), because y_t is dependent on y_{t-2} through y_{t-1} . Suppose we break this chain of dependence by removing ("partialing out") the effect y_{t-1} . Then, we consider the correlation between $[y_t - \phi_1 y_{t-1}] \& [y_{t-2} - \phi_1 y_{t-1}] - i.e.$, the correlation between $y_t \& y_{t-2}$ with the linear dependence of each on y_{t-1} removed: $\gamma(2) = Cov(y_t - \phi_1 y_{t-1}, y_{t-2} - \phi_1 y_{t-1}) = Cov(\varepsilon_t, y_{t-2} - \phi_1 y_{t-1}) = 0$ Similarly, $\gamma(k) = Cov(\varepsilon_t, y_{t-k} - \phi_1 y_{t-1}) = 0$ for all k > 1.

Partial ACF

Definition: The **PACF** of a stationary time series $\{y_t\}$ is ϕ_{hh} : $\phi_{11} = \operatorname{Corr}(y_t, y_{t-1}) = \rho(1)$ $\phi_{hh} = \operatorname{Corr}(y_t - \operatorname{E}[y_t | I_{t-1}], y_{t-h} - \operatorname{E}[y_{t-h} | I_{t-1}])$ for h = 2, 3, ...This removes the linear effects of $y_{t-2}, ..., y_{t-h}$. **Example:** AR(p) process: $y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \varepsilon_t$ $E[y_t | I_{t-1}] = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_h y_{t-h-1}$ $E[y_{t-h} | I_{t-1}] = \mu + \phi_1 y_{t-h-1} + \phi_2 y_{t-h-2} + ... + \phi_h y_{t-h}$. Then, $\phi_{hh} = \phi_h$ if $1 \le h \le p$ = 0 otherwise \Rightarrow After the p^{th} PACF, all remaining PACF are 0 for AR(p) processes.

Partial ACF

• The PACF ϕ_{hh} is also the last coefficient in the **best linear** prediction of y_t given $y_{t-1}, y_{t-2}, ..., y_{t-h}$. (\Rightarrow OLS!)

OLS estimation steps:

Regress y_t against $y_{t-1} \Rightarrow \phi_{11}$: estimated coefficient of y_{t-1} .

Regress y_t against $y_{t-1} \& y_{t-2} \Rightarrow \phi_{22}$: estimated coefficient of y_{t-2} .

Regress y_t against $y_{t-1}, y_{t-2}, \dots, y_{t-h} \Rightarrow \phi_{hh}$: estimated coefficient of y_{t-h} .

• OLS estimation is simple, easy to use. Estimation by Yule-Walker equation is possible. The is also a recursive algorithm by Durbin-Levinson.

• The plot of the PACF is called the **partial correlogram**.







<u>Note</u>: The PACF can be calculated by h regressions, each one with h lags. The hh coefficient is the hth order PACF. Using *ar* R function:

```
> ar(sim_ar2, order.max=1, method = "ols")

Coefficients:

1

0.5586 \Rightarrow \phi_{11} = 0.5586

Intercept: -0.008403 (0.0761)

Order selected 1 sigma^2 estimated as 1.152

> ar(sim_ar2, order.max=2, method = "ols")

Coefficients:

1 2

0.3974 0.2869 \Rightarrow \phi_{22} = 0.2869

Intercept: -0.009847 (0.07326)

Order selected 2 sigma^2 estimated as 1.063
```

Partial ACF – MA(q)

• Following the analogy that PACF for AR processes behaves like an ACF for MA processes, we will see exponential decay (*"tails off"*) in the partial correlogram for MA process. Similar pattern will also occur for ARMA(p, q) process.

Example: We simulate an MA(1) process with $\theta_1 = 0.5$. sim_ma1 <- arima.sim(list(order=c(0,0,1), ma = 0.5), n=200) > pacf(sim_ma1)





PACF – Example: U.S. Stock Returns **Example: US Monthly Returns** (1871 – 2020, *T*=1,795) $pacf_p \le acf(lr_p)$ # pacf: R function that estimates the PACF > pacf_p Partial autocorrelations of series 'lr_p', by lag 0.278 -0.081 -0.026 0.041 0.058 0.002 0.038 0.032 0.016 0.022 0.009 -0.023 -0.057 -0.032 -0.045 0.027 0.017 0.037 -0.059 -0.051 -0.050 0.005 24 $0.006 \ 0.004 \ -0.005 \ -0.051 \ 0.014 \ -0.007 \ 0.037 \ 0.008 \ 0.018 \ 0.023$ $SE(r_k) = 1/sqrt(1,795) = .0236.$ \Rightarrow 95% C.I.: \pm 2* 0.0236





ARIMA Models: Identification – Correlations

• Correlation approach.

Basic tools: sample ACF and sample PACF.

- ACF identifies order of MA: Non-zero at lag q; zero for lags > q.
- PACF identifies order of AR: Non-zero at lag *p*; zero for lags >*p*.
- All other cases, try ARMA(p, q) with p > 0 and q > 0.

<u>Summary</u>: For p > 0 and q > 0.

	AR(p)	MA(q)	ARMA(<i>p</i> , <i>q</i>)
ACF	Tails off	0 after lag q	Tails off
PACF	0 after lag p	Tails off	Tails off

<u>Note</u>: Ideally, "Tails off" is exponential decay. In practice, in these cases, we may see a lot of non-zero values for the ACF and PACF.